Department of Mathematics
College of William & Mary

MATH 451/551

Chapter 4. Common Discrete Distributions

4.6 Hypergeometric Distribution

GuanNan Wang
gwang01@wm.edu

Notes

- The binomial distribution can be applied to probability problems involving sampling n items with replacement from a urn containing two different types of items.
- One type of item is generically referred to as “success" and the other type of item is generically referred to as “failure".
- What happens if the sampling is performed without replacement?

Motivating Example

12 eggs $\left\langle \begin{array}{l} 9 \text{ good.} \\ 3 \text{ bad} \end{array} \right.$

If a dozen (12) eggs contains three bad eggs, find the probability of obtaining exactly x defective eggs in a sample of size 5.

Let X = # bad eggs in a sample of size 5.

① with replacement

$$\left\{ \begin{array}{l} n = 5 \\ \left\langle \begin{array}{l} \text{success} - \text{a bad egg} \\ \text{fail} - \text{a good egg.} \end{array} \right. \\ \text{indep.} \\ P(\text{success}) = P(\text{a bad egg}) = \frac{3}{12} = \frac{1}{4} \end{array} \right\} X \sim Bin\left(n=5, p=\frac{1}{4}\right)$$

$$f_X(x) = \binom{n}{x}\left(\frac{1}{4}\right)^x\left(1-\frac{1}{4}\right)^{n-x}, \quad x = 0, 1, 2, 3, 4, 5$$

② without replacement

X = # bad egg in a sample of size 5.

$n = 5$

- success - a bad egg
- fail - a good egg.

$\frac{3}{12}$ (1st) $\leftarrow$ $\frac{2}{11}$ (2nd). NOT indep

$\frac{9}{12}$ (1st) $\leftarrow$ $\frac{3}{11}$ (2nd)

$P(X=4) = 0$

$X \not\sim \text{Bin}$

$$P(X=0) = \frac{\binom{9}{5}}{\binom{12}{5}}$$

$$P(X=1) = \frac{\binom{3}{1} \times \binom{9}{4}}{\binom{12}{5}}$$

$$P(X=2) = \frac{\binom{3}{2} \times \binom{9}{3}}{\binom{12}{5}}$$

$$P(X=3) = \frac{\binom{3}{3} \times \binom{9}{2}}{\binom{12}{5}}$$

Framework as an Urn Model

Framework as an urn model:

- urn contains N balls
- m of which are "successes"
- $N - m$ of which are "failures"
- random sample of n balls is drawn without replacement

HyperGeometric Distribution

- A discrete random variable X with probability mass function

$$f(x) = \frac{\boxed{\binom{m}{x}}\boxed{\binom{N-m}{n-x}}}{\boxed{\binom{N}{n}}}, \qquad x = 0, 1, 2, \ldots, n$$

for some nonnegative integer parameter N, $n = 0, 1, 2, \ldots, N$, and $m = 0, 1, 2, \ldots, N$, is a *HyperGeometric*(m, N, n) random variable.

$X \sim HyperGeometric(m, N, n)$

$X \sim HG(m, N, n)$

- **Alternative Support:**

$$\mathcal{A} = \{x | x = \max\{0, (m+n-N)\}, \ldots, \min\{n, m\}\}.$$

HyperGeometric Distribution

- **Rationale:**
  - upper bound $\min\{n, m\}$
    - you can't get more successes than n in the sample of size n
    - you can't get more successes than the number of successes m in the population
  - lower bound $max\{0, (m + n - N)\}$
    - you can't get fewer successes than 0 in the sample
    - you are guaranteed to get at least $m + n - N$ successes if the population size is small enough

Mean

$X \sim HG(m, N, n)$

$$f(x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$$

$$E(X) = \sum_x x \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}} = \sum_x \frac{\cancel{x}\,\frac{m!}{\cancel{x!}\,(x-1)!\,(m-x)!}\cdot\frac{(N-m)!}{(n-x)!\{(N-m)-(n-x)\}!}}{\frac{N!}{n!(N-n)!}}$$

$$= \sum \frac{\frac{m\,(m-1)!}{(x-1)!\{(m-1)-(x-1)\}!}\cdot\frac{\{(N-1)-(m-1)\}!}{\{(n-1)-(x-1)\}!\,[\{(N-1)-(m-1)\}-\{(n-1)-(x-1)\}]!}}{\frac{(N-1)!\,N}{(n-1)!\{(N-1)-(n-1)\}!\,n}}$$

$$= \sum \frac{\binom{m-1}{x-1}\binom{(N-1)-(m-1)}{(n-1)-(x-1)}}{\binom{N-1}{n-1}}\cdot\frac{m}{\frac{N}{n}} = \frac{mn}{N}$$

large. N $\quad \frac{m}{N} \to p$

Variance

?

$$V(X) = E(X^2) - \{E(X)\}^2 = E\{X(X-1)\} + E(X) - \{E(X)\}^2$$

$$E\{X(X-1)\} = \sum \cancel{x(x-1)} \frac{\dfrac{m!}{\cancel{x!}(x-2)!\,(m-x)!}\;\dfrac{(N-m)!}{(n-x)!\,\{(N-m)-(n-x)\}!}}{\dfrac{N!}{n!\,(N-n)!}}$$

$$= \sum \frac{\dfrac{(m-2)!\,\boxed{m(m-1)}}{(x-2)!\,\{(m-2)-(x-2)\}!}\;\dfrac{\{(N-2)-(m-2)\}!}{\{(n-2)-(x-2)\}!\,[\{(N-2)-(m-2)\}-\{(n-2)-(x-2)\}]!}}{\dfrac{(N-2)!\,\boxed{N(N-1)}}{(n-2)!\,\boxed{n(n-1)}\,(N-n)!}}$$

$$= \frac{m(m-1)n(n-1)}{N(N-1)} \underbrace{\sum \frac{\binom{m-2}{x-2}\binom{(N-2)-(m-2)}{(n-2)-(x-2)}}{\binom{N-2}{n-2}}}_{1}$$

$$V(X) = \frac{m(m-1)n(n-1)}{N(N-1)} + \frac{mn}{N} - \frac{m^2n^2}{N^2} = \frac{mn(N-m)(N-n)}{N^2(N-1)}$$

R Functions

R Functions

Function	Returned Value
dhyper(x, m, n, k)	calculates the probability mass function $f(x)$
phyper(x, m, n, k)	calculates the cumulative distribution function $F(x)$
qhyper(u, m, n, k)	calculates the percentile (quantile) $F^{-1}(u)$
rhyper(m, m, n, k)	generates m random variates

Example 1

5 cards without replacement

A five-card is dealt from a well-shuffled 52-card deck. Let X be the number of diamonds in the hand. Find $P(X = 3)$ and $P(X \leq 3)$.

13

X: $\mathcal{A} = \{0, 1, 2, 3, 4, 5\}$ $m = 13$. $N = 52$ $n = 5$

$$f(x) = \frac{\binom{13}{x}\binom{52-13}{5-x}}{\binom{52}{5}} = \frac{\binom{13}{x}\binom{39}{5-x}}{\binom{52}{5}}$$

$$P(X=3) = \frac{\binom{13}{3}\binom{39}{2}}{\binom{52}{5}} = \text{dhyper}(x=3, m=13, \underbrace{52-13}_{39}, 5)$$

$$P(X \leq 3) = \sum_{x=0}^{3} \frac{\binom{13}{x}\binom{39}{5-x}}{\binom{52}{5}} = \text{phyper}(x=3, m=13, 52-13, 5) \approx 0.9888$$

Example 2

A biologist uses a “catch and release” or “mark and recapture" program to estimate the population size of a particular animal in a region. During the catch phase, 20 animals are captured, tagged and released. Several months later, 30 animals are captured, and 7 of them have tags. What is the most likely population size?

X = # tagged animals in the second capture

$X \sim HG(m=20, \boxed{N}, n=30)$

$f(7) = \underbrace{P(X=7)}_{\text{maximized}} = \dfrac{\binom{20}{7}\binom{N-20}{30-7}}{\binom{N}{30}}$

dhyper(7, m=20, N-20, 30)

$N = 85$

80	81	82	83	84	85	86	87	88
0.2043	0.2065	0.2082	0.2094	0.2101	0.2104	0.2103	0.2098	0.2090

THANK YOU!